

EL GENOMA HUMANO. UN SISTEMA ALTAMENTE COMPLEJO

María Luisa Martínez-Frías.

Departamento de Farmacología, Facultad de Medicina, Universidad Complutense. Madrid.

Centro de Investigación sobre Anomalías Congénitas (CIAC), Instituto de Salud Carlos III, Ministerio de Sanidad y Consumo. Madrid.

Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Madrid.

Summary

Title: The Human Genome. An Extremely Complex System.

Since the completion of the sequence of the human genome, knowledge of the structure and function of DNA is growing dramatically. However, at the same time, studies are showing an impressive complexity in the structural and functional aspects of DNA. One of the first findings was the identification of frequent DNA variants known as *Single Nucleotide Polymorphisms* (SNP). More recently, and particularly since 2004, a high degree of fine scale structural variations in the human genome are being identified, which at present are globally called *Copy Number Variations/variants* (CNVs). As these CNVs are quite frequent, they have been considered polymorphisms, and are responsible for a greater individual variability than SNPs, with a genetic variation estimated in 1/800 bp, while the SNPs vary in 1/1,200 bp. The CNVs are classified into those altering the number of DNA copies, such as insertions, deletions, and duplications, and those affecting its position (translocations) or its orientation (inversions). In addition, according to the number of repeat copies and, therefore, size, CNVs are classified as (Table 2): *Large-scale Copy Variations* (LCV), *Intermediate-sized Structural Variants* (ISV) and *Low Copy Repeats* (LCR).

Although there are evidences that the CNVs, even including genes, do not necessarily have adverse effects on individuals who carry them, they may have adverse consequences even when they only include non-coding DNA (ncDNA). Some of them could alter meiotic chromosome pairing giving rise to gametes carrying unbalanced chromosome constitutions. In addition, there are several publications showing that some CNVs are related with malformations and syndromes (holoprosencephaly¹⁹, Peters anomaly²⁴, Townes-Bröcks syndrome²⁵, Cleidocranial dysplasia¹⁹, Campomelic dysplasia²¹⁻²³, and other skeletal dysplasias²⁶), that are due to position-effect and other types of effects such as alteration of gene dosage or the presence of unbalanced chromosomal alterations. In 2006, Redon *et al.*⁹, published a map with the global variation in copy numbers in each of the 46 chromosomes of the human genome.

These results parallel the identification of new transcriptional processes, that have also increased since the recognition that the number of genes in the human genome is slightly more than 1/4 of previous estimates. Several studies have recently shown that nearly the whole human genome is transcribed, and that about 98% of the human genome that is transcribed represents non-coding RNA (ncRNA). This has led to many questions regarding their functional meaning, its relationship with RNA coding proteins, and its implications in the regulation and structural organization of the genome. But, at the same time, there has been an increasing knowledge on the function of different ncRNAs that shows an intricate pattern of interrelations and imbrications. Functionally, these ncRNAs are separated into two groups. One includes the housekeeping ncRNAs that is necessary for the normal function of the cells, such as RNA of transference, nuclear RNAs, ribosomal RNAs, etc. The other group includes the ncRNAs regulators that are expressed in embryonic development during cell differentiation, or as response to different stimuli, and can affect the expression of other genes. Among these are the riboswitches and others that participate in regulating gene expression and transcription and post-transcription processes such as microRNA (miRNA) and interference RNA (iRNA). Recent studies have observed that miRNA can use the interference pathway to activate genes, which is a surprising finding⁴¹⁻⁴². Structural and functional investigations on the different ncRNAs have shown that several of them are related with some human diseases and defects (Table 4). In addition, this year, studies on introns, which are a source of miRNA in different animals (*D. melanogaster*, *C. elegans*), have identified a different class of miRNA precursors, called "mirtrons" whose function is yet unknown. However, recent studies have suggested that they may function in the regulatory biological network, and that they may also exist in other species⁴³⁻⁴⁴.

The recent publication of the results of the pilot study of the ENCODE Project (**ENCyclopedia Of DNA Elements**)²⁷⁻²⁸, has offered a highly complex structural and functional view of the human genome and in the structure of RNAs³⁶, as well as the implications in the alternative transcription (splicing). In relation with this last process, the current results on the function of the alternative proteins suggest that it may not be exactly as previously considered³⁷.

All these findings have led to the revision of previous concepts, starting by the "dogma" of the gene definition. The proposed definition, which represents a good example of the complexity of the human genome, is as follows: **The gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products.**³⁰

This article presents a simplified general review of the most recent findings summarized above, reviewing: a) the structural variations of DNA and functional elements of the human genome, and b) the types and function of different RNAs.

Finally, as a reflection, it is clear that the DNA code is much more than the lecture of the combination of the four bases (adenine, guanine, thymine, and cytosine), and although our knowledge of their structure and function it is still very small, its high level of complexity is becoming increasingly evident. New information shows multiple and complex frameworks of different functional networks, whose products are not the sum of their components, but "emergent behaviors" in relationship with the other parts of the whole system, although its laws are still unknown. All these aspects are concordant with the characteristics of the so called "*complex systems or chaotic systems*". These, as the chaos theory postulates, do not have absence of order or causality, but a particular interrelationship that gives rise to new levels that are subject to their own emergent rules. Thus, it is possible that the rules that command the human genome are not physically or mathematically different from those that conduct the "Complex systems" regulating nature.

Introducción

Muy pocos, en la actualidad, incluso sin tener conocimientos muy profundos sobre biología, dudan de la enorme importancia que ha tenido el descubrimiento de la estructura del ADN, seguido por el desciframiento del código genético del ser humano, y mucho más recientemente por la secuenciación de ese código. Este último paso está dando lugar a un impresionante aporte de conocimientos sobre las claves de su estructura y funcionamiento, y por ende, de la investigación sobre sus implicaciones en la aparición de numerosas patologías, tanto de las muy frecuentes como de las que son de muy baja frecuencia (hoy llamadas enfermedades raras). Es más, en los últimos tiempos se vienen produciendo hallazgos importantes, y en muchos casos inesperados, sobre los elementos estructurales y funcionales del genoma humano. Cada nuevo paso en esta investigación nos muestra que la mayoría de los aspectos que considerábamos como hechos claros, en realidad representan una esquemática simplificación de procesos mucho más complejos. Tanto, que están obligando a revisar la mayoría de las definiciones y significados plenamente establecidos desde hace décadas, incluyendo el que ha sido el "dogma" principal de los procesos biológicos: el propio concepto de lo que es un gen.

Uno de los primeros indicios de que no todo era tan sencillo en la función de los genes en el ser humano, fue la constatación de que ciertos genes no producen el mismo efecto independientemente del progenitor del que se reciben, sino que tienen manifestaciones diferentes según si proceden de la madre o del padre (procesos de "imprinting" génico). Estos procesos epigenéticos son modificaciones reversibles de la función sin alterar la secuencia estructural del ADN, mediante la metilación/desmetilación de la citosina, o por modificación de las histonas. La metilación del ADN tiene un efecto de silenciamiento de la actividad del gen, y la desmetilación la activación funcional del mismo. De hecho, durante las primeras fases del desarrollo embrionario, todo el genoma sufre una desmetilación general para alcanzar la pluripotencialidad, que irá seguida de una nueva metilación selectiva durante los procesos de diferenciación celular, y el establecimiento y desarrollo de los primordios de los diferentes órganos y estructuras. Por otra parte, en estudios recientes se han identificado nuevas funciones de la metilación de las histonas y organización de la cromatina, en la expresión y función del genoma.

Esos hallazgos fueron seguidos por otros muchos en relación no sólo con las alteraciones crípticas de los cromosomas, sino especialmente con la estructura del ADN, organización de los genes y de sus productos y, en este mismo año, sobre los elementos funcionales del genoma. Resultados

que están siendo espectaculares, en gran parte por el desarrollo de las tecnologías molecular y bioinformática. Esta última es imprescindible para que se pueda analizar la gran cantidad de información obtenida, así como para el estudio "in silico" (mediante análisis con ordenadores) de simuladores dinámicos moleculares dirigidos al estudio de los mecanismos por los que se produce la función. La consecuencia es que desde hace cuatro o cinco años, se viene produciendo una auténtica revolución en los análisis a gran escala, que han devenido en otros muchos niveles de estudio, por lo que a la investigación del genoma se han ido añadiendo los denominados *proteoma*, *metiloma*, *RNoma*, *transcriptoma*, *interactoma*... El resultado es la constatación diaria de que nos encontramos ante una complejísima estructura. Complejidad que afecta tanto a su organización y su función, como al gran entramado de interrelaciones, superposiciones, interacciones estructurales y funcionales, acompañada de una alta variabilidad individual, sobre las que nuestros conocimientos son aún rudimentarios. Incluso se ha mostrado que ciertas variaciones estructurales del genoma predisponen a la aparición de reordenamientos nuevos, que pueden favorecer la aparición de enfermedades, defectos congénitos y diversos tipos de cáncer.

Sería, pues, fatuo por mi parte intentar siquiera resumir todos estos avances en unos pocos folios, y en una forma suficientemente clara para los lectores que no estén muy familiarizados con esta área de la investigación biológica. Por ello, sólo voy a describir someramente un panorama muy global de la situación actual, incluyendo algunas definiciones de los conceptos y aspectos más novedosos. El objetivo es sólo que los lectores del *Boletín del ECEMC*, puedan tener una idea general sobre los derroteros por los que avanza la investigación sobre el genoma humano, junto con el espectacular incremento de conocimientos que se está generando en los aspectos estructurales y funcionales del mismo. Para ello, y en aras de la claridad, vamos a resumir en primer lugar algunos de los resultados obtenidos en relación con las variaciones de la estructura del ADN, y elementos funcionales del genoma humano y, en segundo lugar, los relacionados con la estructura y función de los distintos ARNs.

Variaciones estructurales del ADN y elementos funcionales del genoma humano

Variaciones estructurales del ADN

Hasta hace unos pocos años las alteraciones estructurales que se conocían eran aquellas que podían visualizarse con el microscopio en preparaciones citogenéticas con cro-

mosomas de alta resolución (en las que se consiguen unas 850 bandas por cariotipo). Es decir, alteraciones de la estructura del ADN con tamaños superiores a 3-7 Mb (Tabla 1). No obstante, también se detectaban alteraciones más pequeñas mediante técnicas de hibridación *in situ* con fluorescencia (FISH, de sus siglas en inglés), denominadas microdelecciones, que afectan a diferentes regiones de los cromosomas incluyendo las regiones subteloméricas. Incluso se habían detectado otras alteraciones menores aún mediante técnicas moleculares, consistentes en secuencias duplicadas (como en el síndrome de Williams). Muchos de estos casos, por sus manifestaciones clínicas habían sido previamente considerados como pacientes con síndromes de genes contiguos (Smith-Magenis, Prader-Willi/Angelman, entre otros).

TABLA 1

DEFINICIÓN Y ABREVIATURAS DE TAMAÑOS DE ADN

DEFINICIÓN DEL TAMAÑO	ABREVIATURAS
Un par de bases	pb
10 ³ pares de bases (kilo-bases)	kb
10 ⁶ pares de bases (mega-bases)	Mb

Sin embargo, una demostración de la estrecha relación existente entre el desarrollo de la tecnología, la investigación biológica y el enfoque multidisciplinar, es la gran ayuda que ha supuesto la utilización de la bioinformática. Así se pudo determinar la frecuencia de los primeros polimorfismos que se identificaron, los denominados cambios de un solo nucleótido (*Single nucleotide polymorphism-*

SNP), consistentes en la sustitución de una base (Tabla 2). Estos han mostrado ser muy frecuentes en la población humana, habiéndose estimado que existen entre 10 y 15 millones de SNPs en nuestro genoma¹, algunos de los cuales se han relacionado con problemas de susceptibilidad.

Además, en sólo unos pocos años, se viene produciendo un ingente aporte de información biológica derivada del análisis de los hallazgos moleculares sobre otros muchos aspectos del genoma humano, mediante tratamientos bioinformáticos. En el año 2004, se publicaron dos trabajos²⁻³ en los que se mostraba que nuestro genoma tiene una gran cantidad de variaciones estructurales submicroscópicas, consistentes en grandes fragmentos de ADN duplicados o perdidos. Estas variaciones se observan en más del 1% de la población, por lo que se consideran polimórficas, y han dado lugar a una nueva, e inesperada, visión sobre la estructura y función del genoma humano. En sólo tres años transcurridos desde esas dos publicaciones, se ha podido determinar que las diferencias genéticas entre dos personas no se encuentran sólo a nivel de la secuencia de bases del ADN, sino fundamentalmente en las variaciones estructurales del mismo. Por ello, en Julio de este mismo año se ha publicado un número especial (vol S39) de la revista *Nature Genetics*, bajo el título de "Variaciones estructurales del genoma". Éstas han recibido diferentes denominaciones en la literatura científica, por lo que en la Tabla 2 se resumen las definiciones más generalizadas en este momento, y los distintos grupos en los que se han separado.

Según se observa en la Tabla 2, estas variaciones estructurales polimórficas consisten en secuencias repetidas del ADN, de número variable de repeticiones, que tienen tamaños mayores que los SNPs, pero, en general, menores de

FIGURA 1. Esquema de los diferentes tipos de CNVs.

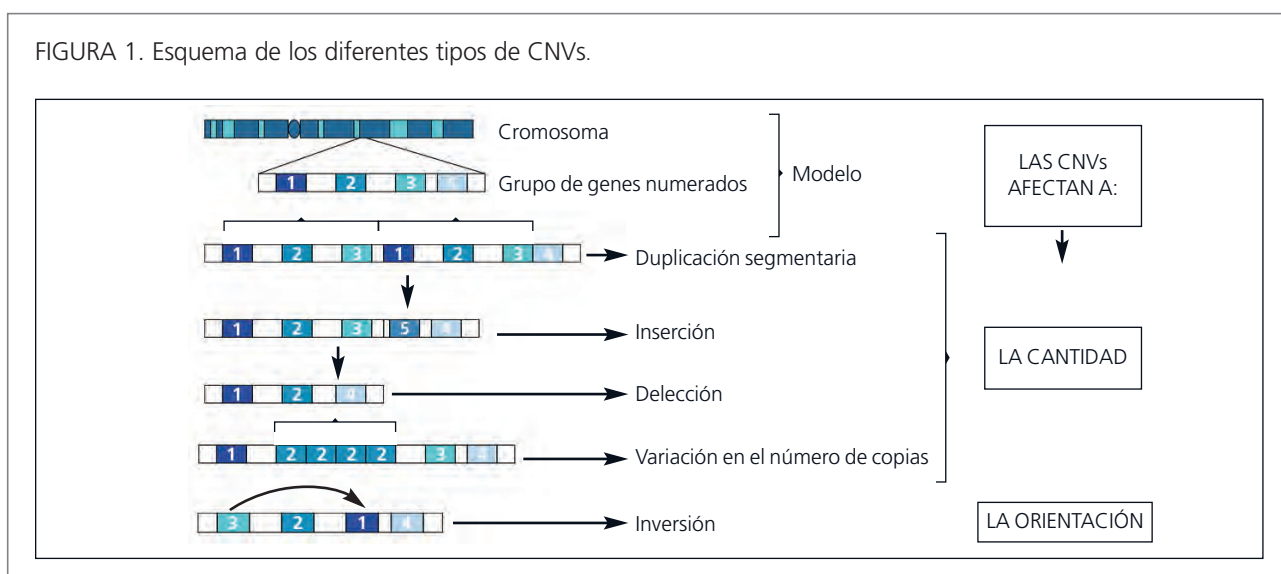
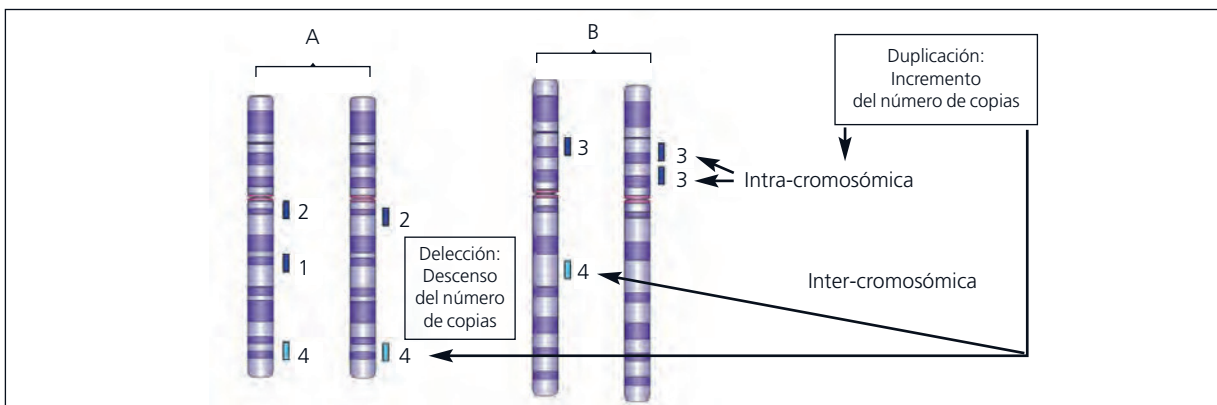


TABLA 2

TIPOS DE VARIACIONES POLIMÓRFICAS DEL GENOMA HUMANO

TIPOS	DESCRIPCION
Polimorfismos de un solo nucleótido (SNP):	Consisten en la sustitución en un solo nucleótido. Se considera que hay alrededor de 10-15 millones en el genoma humano, con un promedio de diferencia en un SNP por cada 1.250 pb entre individuos elegidos al azar.
Variantes estructurales Globalmente llamadas CNVs (<i>Copy number variants-CNV</i>). (Incluyen todos los cambios que no son sustituciones de una sola base)	En amplio sentido se refieren a alteraciones estructurales polimórficas que afectan a segmentos del ADN mayores de 1kb. Son de diferentes tipos: Cuantitativas, de posición y de orientación. Se las llama globalmente como "Variantes de número de copias (CNVs)". Tienen un promedio de diferencia en un SNP por cada 800 pb entre individuos elegidos al azar.
a) Variantes cuantitativas	Incluyen variaciones en el número de copias. Son polimórficas cuando se encuentran en más del 1% de la población. Sus tipos son inserciones, duplicaciones, y deleciones. A veces se agrupan formando regiones de CNVs, llamadas CNVR. Se han separado, por su tamaño en varios tipos, siendo los principales los siguientes:
Inserción/delección (Indel)	Se refieren a la pérdida/ganancia de un pequeño segmento (menor de 1kb) en la secuencia del genoma (uno o varios nucleótidos).
Variantes de pocas copias Repetidas (<i>Low copy repeat-LCR</i>), o Duplicones	Se refieren a las variantes de un tamaño inferior a 1kb de ADN, y el número de copias varía entre distintos individuos. También se llaman duplicaciones segmentarias (<i>Segmental duplication-SD</i>).
Variantes estructurales de tamaño intermedio (<i>Intermediate-sized structural variants-ISV</i>).	Son variantes estructurales de secuencias de ADN de tamaños entre 8 y 50 kb.
Variantes de gran tamaño de número de copias (<i>Large-scale Copy variation LCV</i>).	Son variaciones estructurales de gran tamaño (>50 kb).
b) Variantes posicionales y de orientación	Son re-arreglos estructurales muy pequeños, como inversiones o translocaciones.

FIGURA 2. Dos parejas de cromosomas homólogos.



Los bloques numerados a la derecha de los cromosomas de las parejas A y B, indican la localización de CNVs. El bloque número 1, está delecionado en uno de los cromosomas A. El bloque numero 3, está duplicado en la pareja B, por lo que ambas CNVs son intra-cromosómicas. El bloque 4 de la pareja A, está duplicado pero en otro cromosoma no homólogo, el B, por lo que es una CNV inter-cromosómica.

TABLA 3

**SÍNDROMES CON DEFECTOS CONGÉNITOS CUYAS ALTERACIONES GENÓMICAS ESTÁN MEDIADAS
POR DUPLICACIONES SEGMENTARIAS ***

SÍNDROMES	ALTERACIÓN	TAMAÑO EN Mb	LOCALIZACIÓN
Smith-Magenis	Delección	5	17p11.2
Neurofibromatosis 1	Delección	1,5	17q11.2
Prader-Willi/ Angelman	Delección	4	15q11-13
Williams-Beuren	Delección	1,6	7q11.23
DiGeorge/ Velocardiofacial	Delección	3	22q11.2
Ictiosis ligada al sexo	Delección	1,9	Xp22
Marcadores cromosómicos extras	Duplicación invertida	4	15q11-14
" Cat eye " (marcador)	Extra marcador	3	22q11.2
Retraso mental y rasgos dismórficos	Microdelección	<1	17q21.31

* Emmanuel and Shaikh Nature Review 2001;2:791-800. Modificado.

3Mb, por lo que no son visibles en estudios citogenéticos²⁻⁴. No obstante, hay algunos casos en los que se han descrito variaciones visibles por citogenética de alta resolución, que no se relacionaban con alteraciones fenotípicas y que se habían denominado "variantes eucromáticas"⁵.

En amplio sentido y de una forma global, las variaciones estructurales se denominan como "Variaciones/variantes de número de copias" (*Copy Number Variations/variants-CNVs*), aunque han recibido otras denominaciones dependiendo del tipo de alteración estructural y su tamaño (Tabla 2)^{2-3,6}. Las CNVs se han separado en dos grupos: las que implican alteración del número de copias (inserciones, duplicaciones, o delecciones), y las que afectan a la posición (translocaciones), o a la orientación (inversiones) (Figura 1). Las CNVs pueden ser inter-cromosómicas, en las que los segmentos repetidos del ADN están distribuidos entre cromosomas no homólogos, o intra-cromosómicas, cuando se encuentran en el mismo cromosoma (Figura 2). Las CNVs, producen mayor variabilidad individual que los SNPs, ya que implican una variación genética de 1/800 pb entre individuos, mientras que los SNPs representan 1/1.200 pb⁷.

Aunque hay evidencias de que grandes CNVs, incluso conteniendo genes, no necesariamente dan lugar a alteraciones, o enfermedades en los individuos que las tienen, también pueden producir problemas incluso cuando sólo incluyen segmentos de ADN no codificante (ADNnc). De hecho, pueden dar lugar a re-arreglos cromosómicos diferentes durante el proceso de meiosis, algunos de los cuales podrían ser deletéreos. Por tanto, cabe la posibilidad de que ciertas variantes puedan ser polimórficas en un grupo, y producir una enfermedad (o defecto) en otro. Algún estudio ha mostrado que, aunque las CNVs se distribuyen por todo el genoma, en ciertos lugares se producen agrupamientos de unas pocas, que constituyen zonas "calientes" para re-arre-

glos cromosómicos, tanto para variaciones normales, como también para las que producen enfermedades^{2,8-18}. Así se ha observado para algunos cuadros clínicos bien conocidos, incluyendo varios con defectos congénitos debidos a microdelecciones (Tabla 3), en los que se han encontrado CNVs en las regiones cromosómicas en las que se localiza la delección, indicando que inducen inestabilidad de esas regiones. Recientemente, tres grupos diferentes¹⁶⁻¹⁸, han identificado un nuevo síndrome de microdelección en la región q21.31 del cromosoma 17 (17q21.31), en pacientes que presentaban retraso mental, hipotonía, alteración (o ausencia) del lenguaje y de la marcha, junto a rasgos dismórficos. Estos autores detectaron que dicha región del cromosoma 17, estaba flanqueada por duplicaciones segmentarias, cuyos re-arreglos daban lugar a la microdelección.

Las CNVs, también pueden producir otros problemas por diferentes mecanismos, como variación de la dosis génica a través de la interrupción directa de genes, por unión de genes, por haploidía de un gen recesivo, o por un "efecto de posición". Este último consiste en la alteración del patrón de expresión de un gen como resultado de su localización genómica, o de la cromatina que lo rodea, y puede producirse por varios mecanismos. Por ejemplo, la transcripción de ciertos genes a una posición que está muy cercana a elementos reguladores positivos, o por separación/agrupación de genes de sus elementos de control de transcripción, ya que elementos reguladores de la expresión de ciertos genes, se pueden encontrar separados de ellos a distancias de varios millones de pares de bases⁶. Hay algunos trabajos¹⁹⁻²⁵ que sirvieron de base para establecer algunos de los mecanismos antes mencionados, que nos permiten documentar la importancia que tienen las CNVs en las alteraciones del desarrollo embrionario y fetal, dando lugar a defectos congénitos. Por ejemplo, David y col.²⁴, observaron un

paciente con anomalía de Peters (OMIM 604229) -un defecto de la cámara anterior del ojo- que tenía una translocación entre los cromosomas 1 y 7. En el cromosoma 7p21.1 se rompía el gen de la histona deacetilasa (*HDAC9*), mientras que la rotura recíproca en el cromosoma 1 se encontraba a una distancia aproximada de 500 kb 3' del gen del factor de crecimiento de transformación beta 2 (*TGFB2*). Como en el ratón el gen *Tgfb* produce un defecto ocular muy similar al de Peters, se consideró que la anomalía del paciente se debía a un efecto de posición del gen *TGFB2*, más que a la rotura del gen *HDAC9*. Otro ejemplo²⁵ es el de un paciente con diagnóstico de Síndrome de Townes-Bröcks (OMIM 107480), que tenía una translocación balanceada entre los cromosomas 5 y 16. La rotura del cromosoma 16q estaba a una distancia de 180 kb 5' de la localización del gen *SALL1* responsable de este síndrome, que está localizado en 16q12.1, por lo que también se consideró resultado de un efecto de posición²⁵.

En el año 2006, Redon et al.⁹, publicaron los resultados del análisis global del genoma para la identificación de todos los tipos de variantes del número de copias, mapa que está siendo de gran utilidad para la investigación funcional y sus efectos. Este mismo año se ha publicado un caso estudiado en el ECEMC²⁶, de una pareja sana pero con una historia previa de abortos por defectos congénitos. De sus cuatro embarazos, tres habían terminado en abortos, dos de ellos porque los fetos tenían enanismo (con cariotipo normal por amniocentesis), y el otro, por un exceso de material en el cromosoma 4p, identificado también por amniocentesis. En el estudio con cromosomas de alta resolución de la pareja, el padre resultó ser portador sano de una alteración cromosómica de tamaño muy grande (3,3 Mb) en la región 4p16.1. El análisis molecular mostró que correspondía a un segmento duplicado, que se encontraba a una distancia de entre 6 y 8 Mb del gen del receptor 3 del factor de crecimiento de fibroblastos (*FGFR3*), que está localizado en 4p16.3, y que da lugar a varios tipos de condrodisplasias (como la acondroplasia y la displasia tanatofórica). En el reciente mapa de las CNVs en el ser humano⁹, se observa la existencia de CNVs en la misma región del brazo corto del cromosoma 4 en la que se identificó la duplicación del caso del ECEMC²⁶. Por tanto, es probable que la alteración estructural del padre sea una variante de gran tamaño de número de copias-LCV (Tabla 2). Esta podría haber producido gametos alterados por las dificultades del apareamiento cromosómico durante la meiosis y la recombinación. Las alteraciones de los gametos serían diferentes, pudiendo producir tanto excesos y defectos de material en 4p, como modificación de la posición de las CNV y un efecto de posición, lo que también podría explicar el enanismo de dos de los fetos.

Elementos funcionales del genoma

Los análisis estructurales y funcionales del genoma humano, se tienen que realizar comparativamente en diferentes grupos de individuos debido tanto a la gran variabilidad observada intra e inter-individuos, como a las diferentes frecuencias de las distintas variantes.

En los primeros estudios ya se confirmó que esa gran cantidad de material genético tradicionalmente considerada como "ADN basura", y luego por algunos como "zona negra" o "zona no codificante" del genoma, tiene una importantísima relevancia en la actividad funcional de los genes (o porciones codificantes de proteínas), incluso distantes entre sí. Esto ha sugerido que las variaciones estructurales que se han identificado en el genoma, podrían relacionarse con los procesos de transcripción y sus inicios, pero también con grandes dominios de activación/desactivación de cromatina, así como con los requerimientos para la replicación del ADN. Es más, los estudios que han analizado la variación genética entre individuos en fracciones pequeñas del genoma, han ido mostrando que esas áreas de material genético "no codificante" (ADNcn) son las que establecen las mayores diferencias entre el hombre y el resto de mamíferos, y no el número de genes, como se había considerado previamente que, además, constituyen un pequeño porcentaje de todo el genoma humano.

Por todo esto, hace unos pocos años se constituyó un Proyecto denominado ENCODE (**ENC**yclopedia **O**f **DNA E**lements)²⁷, mediante un consorcio entre 35 grupos de investigación de diferentes países, con el propósito de identificar los elementos funcionales del genoma humano (www.genome.gov/ENCODE). Es decir, las secuencias de ADN que controlan el funcionamiento de los genes.

Este mismo año, se acaban de publicar los resultados de la fase piloto del Proyecto ENCODE²⁸, siendo algunos de los más importantes los siguientes: a) Que la mayoría del control de funcionamiento de los genes parece que se encuentra en las zonas de ADNnc y, por tanto, en las que no producen una proteína como producto final detectable. b) Que el número de genes (en el sentido clásico de secuencias de ADN que se transcriben al ARN mensajero para producir una proteína), o ADN codificante, representa una parte muy pequeña del ADN humano. c) Que se transcribe la gran mayoría del genoma (si no todo), y no sólo las secuencias que codifican proteínas, como se pensaba hasta ahora. d) Que no todas las secuencias de ADN codificante están conservadas evolutivamente, mientras que muchas de las secuencias del ADNnc tienen papeles reguladores esenciales, y se han mantenido sin cambios a lo largo de la evolución²⁹. Esta observación va en contra de lo establecido previamente, sobre que las secuencias de ADN codificante eran altamente con-

servadas porque su funcionamiento correcto era necesario para la supervivencia.

Todos estos hallazgos han llevado a la necesidad de establecer nuevas definiciones e, incluso, modificar muchos de los conceptos existentes, entre ellos, como ya se comentó al principio de este artículo, el de gen. Pero estos cambios no son fáciles, bien porque la complejidad estructural y funcional del genoma lo impide, o bien por el gran desconocimiento que aún tenemos del mismo. Como ejemplo de esas dificultades, basta con conocer la propuesta que se ha hecho este mismo año para la definición de gen: "**Un gen es una unión de secuencias genómicas que codifican un conjunto coherente de productos funcionales que potencialmente pueden superponerse**"³⁰.

Todo esto, lleva a la conclusión de que nada hay en la estructura del ADN que sea superfluo, carente de función o, simplemente, "basura". Lo que ocurre es que cada parte del mismo, participa en la función y su control, en una forma sobre la que nuestro conocimiento es aún muy pequeño. Y, por lo que se viene observando en los últimos años, mediante unos procesos de extraordinaria complejidad.

Tipos y función de los ARNs

Una de las primeras sorpresas que deparó el desciframiento de nuestro genoma, fue que el número de genes era casi la cuarta parte del que se habían establecido en función al número de proteínas que se conocían. En consecuencia, la concepción tradicional de que la información genética contenida en el ADN se transcribe formando una molécula de ARN mensajero (ARNm) que en el citoplasma celular fabrica una proteína, debía ser una simplificación de un proceso mucho más complicado.

Estudios más recientes han ido constatando la simplificación de nuestros conceptos previos, mediante la identificación de nuevos aspectos de la replicación y transcripción de ADN que, en cierta medida, contribuyen a entender la diferencia entre genes y proteínas³¹⁻³⁴. Entre ellos podemos destacar: los procesos de transcripción alternativa ("*splicing*") y de la remodelación de la cromatina, así como la modificación post-transcripcional, consistente en procesos de fosforilación, acetilación, metilación... de las proteínas, mediante los que regulan el tipo de función de cada una en múltiples y diferentes lugares. Estos procesos coordinan las señales inter e intramoleculares mediante un control cualitativo y cuantitativo de la función de esas proteínas³⁵. No obstante, en el Proyecto ENCODE, que también ha abordado el estudio de la estructura y funcionamiento de los ARNs, y las implicaciones de la transcripción alternativa³⁶, se han obtenido resultados intrigantes. Esos resultados sugieren

que aunque el proceso de "*splicing*" parece ser más frecuente de lo que se había considerado, no puede explicar el total de proteínas convencionales tanto enzimáticas como estructurales. Incluso los autores indican que a pesar de que el "*splicing*" tiene el efecto claro de aumentar la variedad de las funciones génicas, esto no ha sido demostrado a nivel de la proteína³⁷, por lo que en realidad no pasa de ser una hipótesis. También sugieren que el proceso de *splicing* alternativo puede dar lugar a un amplio rango de resultados, muchos de los cuales pueden tener una función con efectos adversos³⁷.

Por otra parte, hace apenas unos pocos meses se ha estimado que alrededor del 98% del genoma humano que se transcribe, representa ARNs que no producen proteínas, por lo que se han considerado transcritores de función desconocida (TFD), o ARN no codificante (ARNnc)^{31, 38-40}. La identificación de la gran porción del genoma que implica los ARNnc, ha llevado a plantear muchas preguntas sobre su significado funcional, su relación con los ARNs que codifican proteínas, y sobre su regulación y organización estructural y genómica.

A pesar de todo esto, se conoce la función en un gran número de ARNnc que, a su vez, muestran un patrón de interrelaciones e imbricaciones de gran complejidad, y se han agrupado en dos clases. Una, que incluye los ARNnc que se podrían llamar de "mantenimiento", que son necesarios para que la función y viabilidad celular sea normal. Entre ellos cabe resaltar los ARN de transferencia, los nucleares pequeños, y los ribosómicos, entre otros. La segunda clase, incluye los que son reguladores (o ribo-reguladores), que se expresan en ciertos momentos del desarrollo durante la diferenciación celular, o como respuesta a estímulos externos, y pueden afectar a la expresión de otros genes. Entre ellos se encuentran los ribo-interruptores (*riboswitches*) que actúan como interruptores génicos⁴⁰, así como otros que participan en funciones de regulación tanto en la transcripción como en la post-transcripción, entre los que se encuentran los microARN (miARN) -que pueden controlar muchos genes- y los ARN de interferencia (ARNi), cuyo descubrimiento supuso el Premio Nobel del año 2006 para los investigadores Craig Mello y Andrew Fire.

Los miARN y ARNi, han abierto un amplio campo de investigación, tanto básica como terapéutica, basada en que los miARN actúan reprimiendo la traducción del ARNm diana o, incluso, destruyendo un ARNm. El miARN primario, una horquilla de alrededor de 100 nucleótidos, se procesa en el núcleo dando lugar al precursor del miARN que es la forma que pasa al citoplasma. En éste, se transforma en miARN mediante la acción de la ribonucleasa *Dicer* (cortadora), y actúa inhibiendo o destruyendo a un ARNm según la complementariedad de sus bases. Sobre los aspectos

TABLA 4

EJEMPLOS DE DEFECTOS CONGÉNITOS RELACIONADOS CON ARNnc

SÍNDROMES	ALTERACIÓN
Displasia campomiélica	(CMPD) ARNnc que altera el nivel de expresión en cáncer
Síndrome de Prader-Willi	(IPW) ARNnc correlacionado con los procesos neurológicos
Síndrome de Prader-Willi	(ZNF127AS) ARNnc correlacionado con los procesos neurológicos
Síndrome de Angelman	(UBE3A-AS) ARNnc correlacionado con los procesos neurológicos
Síndrome de Russell-Silver	(COPG2IT1) ARNnc correlacionado con otros problemas
Síndrome de Russell-Silver	(MESTIT 1) ARNnc correlacionado con otros problemas
Síndrome de DiGeorge	(22k48, HIRA deleción intrón) ARNnc correlacionado con otros problemas
Síndrome de DiGeorge	(DGC55 interrumpido) ARNnc correlacionado con otros problemas
S. Beckwith-Wiedemann	(H19) ARNnc correlacionado con otros problemas
S. Beckwith-Wiedemann	(LIT1) ARNnc correlacionado con otros problemas

funcionales de los miARNs, ya se están haciendo estudios experimentales encaminados a tomar la secuencia de un gen y diseñar un miARN que pueda silenciar de forma específica su ARNm. Sin embargo, estudios recientes han obtenido un hallazgo sorprendente: que usando la vía de la interferencia, los miARNs pueden actuar también activando genes⁴¹⁻⁴², lo que ha planteado nuevos interrogantes sobre la función de estos genes.

Por otra parte, en la investigación sobre los intrones, que son una fuente regular de miARN, realizada en ciertos animales (*D. melanogaster*, *C. elegans*), se ha observado que pueden codificar otra clase diferente de precursores de miARN que se han denominado "mirtrons"⁴³⁻⁴⁴, cuya presencia en la escala evolutiva y su función, aún no es conocida. No obstante, los resultados de esos dos trabajos sugieren que podrían participar en procesos regulatorios, así como que podrían encontrarse en otras especies.

Muchos de los ARNnc descritos en el ser humano, se han relacionado tanto con enfermedades complejas, como con ciertos tipos de defectos congénitos (Tabla 4). Sin embargo, aún queda mucho por conocer sobre estos procesos, que han supuesto una evidencia más de la enmarañada red que constituyen los transcriptores no codificantes de proteínas, como se ha puesto de manifiesto en estudios de este mismo año^{31,36,38,40-42}. De hecho, las técnicas actuales están permitiendo abordar el estudio de la transcripción del genoma mediante análisis globales de lo que se denomina *transcriptoma*.

Parece pues, que la gran cantidad de ARNnc del genoma constituye una complicada red reguladora que actúa en interacción con la red que configuran las proteínas y, aunque aún no se conoce su estructura y función global, los resultados de su investigación están siendo apasionantes.

Aunque cada vez va habiendo más posibilidad para relacionar todos estos resultados con ciertos defectos congénitos (y otros tipos de enfermedades), aún no es posible conocer su función ni sus potenciales efectos. Hasta ahora, los

grupos de población analizados son pequeños, y es necesario disponer de grandes muestras de poblaciones diferentes, para poder analizar la función de esta enorme variabilidad estructural de nuestro genoma, en la que apenas hemos entreabierto el resquicio de una puerta que era desconocida. Por ello, se están desarrollando diferentes bases de datos, muchas disponibles on-line, entre las que podemos citar las siguientes:

[Genomic Variants (www.projects.tcag.ca/variation/).
Human structural variation (www.Humanparalogy.gs.Washington.edu/structuralvariation/).
Developmental genome (www.bwhpathology.org/dgap/).

Algunas consideraciones a modo de reflexión

En estos momentos, es claro que el funcionamiento del código del ADN es mucho más que la simple lectura de las distintas combinaciones de las cuatro bases que lo componen (adenina, guanina, citosina y timina). Y aunque ya es mucho lo que se conoce, representa aún una parte muy pequeña ya que cada paso que se ha dado ha desvelado un nivel de complejidad mucho mayor, mostrando múltiples y complicados entramados de redes estructurales y funcionales, que manifiestan comportamientos nuevos cuyas "leyes" aún son desconocidas. Entre esos comportamientos también podemos identificar adaptabilidad, autoorganización, auto-división, flexibilidad, ... siendo un ingrediente básico la "emergencia", en el sentido de que el comportamiento que manifiesta el conjunto de todas las partes que lo forman, no parece ser el simple resultado de la suma de todas ellas, sino que es diferente (emergente) dependiendo de cuáles y cómo son las otras partes que lo componen en cada caso y momento. Por ello, y aunque son aspectos que han debido ser esenciales en el proceso evolutivo, por

ahora su comportamiento, al menos en cada individuo, es impredecible.

Cualquier lector iniciado en los conceptos y enfoques recientes (aunque no lo sea en los complicados análisis matemáticos) de lo que se denominan "*Sistemas Complejos*, o *Sistemas Caóticos*", reconocerá en las características expuestas en el párrafo anterior, aquellas que definen lo que es un "*Sistema Complejo*". Por ello, las propiedades de complejidad y de emergencia antes mencionadas como rasgos del genoma, no indican ausencia de orden o de causalidad (como determina la Teoría del Caos), sino que implican reglas que dan lugar a nuevos rasgos de mayor nivel que obedecen a su propio tipo de normas emergentes, pero que posiblemente no sean física y matemáticamente diferentes de las que rigen para los "*Sistemas Complejos*" que regulan la propia Naturaleza.

Referencias

- IHMC. A haplotype map of the human genome. *Nature* 2005;437:1299-1320.
- Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Manner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. *Science*. 2004;305:525-528.
- Iafate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nat Genet*. 2004;36:949-951.
- Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurler ME, Carter NP, Scherer SW, Lee C. Copy number variation: new insights in genome diversity. *Genome Res*. 2006;16:949-961.
- Barber JCK. Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. *J Med Genet*. 2005;42:609-629. Review.
- Feuk L, Marshall CR, Wintle RF, Scherer SW. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet*. 2006; Spec1:R57-66. Review.
- Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. Fine-scale structural variation of the human genome. *Nat Genet*. 2005;37:727-732.
- Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Seagraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE. Segmental duplications and copy number variation in the human genome. *Am J Hum Genet*. 2005;77:78-88.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shaper MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurler ME. Global variation in copy number in the human genome. *Nature*. 2006;444:444-454.
- Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurler ME, Dermitzakis ET. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007;315:848-853.
- Lupski JR. Genomic disorders: structural features of genome can lead to DNA rearrangements and human disease traits. *Trends Genet* 1998;14:417-422.
- Lupski JR, Stankiewicz P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet* 2005;1, e49.
- Lupski JR. Structural variation in the human genome. *N Engl J Med*. 2007;356:1169-1171.
- Eichler EE. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet*. 2001;17:661-669. Review.
- Sharp AJ, Selzer RR, Veltman JA, Gimelli S, Gimelli G, Striano P, Coppola A, Regan R, Price SM, Knoers NV, Eis PS, Brunner HG, Hennekam RC, Knight SJ, de Vries BB, Zuffardi O, Eichler EE. Characterization of a recurrent 15q24 microdeletion syndrome. *Hum Mol Genet*. 2007;16:567-572.
- Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, Kooy RF, Reyniers E, Romano C, Fichera M, Schinzel A, Baumer A, Anderlid BM, Schoumans J, Knoers NV, van Kessel AG, Sistermans EA, Veltman JA, Brunner HG, de Vries BB. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet*. 2006;38:999-1001.
- Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, Cumming S, Dunn C, Kalaitzopoulos D, Porter K, Prigmore E, Krepischi-Santos AC, Varela MC, Koiffmann CP, Lees AJ, Rosenberg C, Firth HV, de Silva R, Carter NP. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet*. 2006;38:1032-1037.
- Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC, Fitzpatrick CA, Seagraves R, Richmond TA, Guiver C, Albertson DG, Pinkel D, Eis PS, Schwartz S, Knight SJ, Eichler EE. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet*. 2006;38:1038-1042.
- Fernandez BA, Siegel-Bartelt J, Herbrick JA, Teshima I, Scherer SW. Holoprosencephaly and cleidocranial dysplasia in a patient due to two position-effect mutations: case report and review of the literature. *Clin Genet*. 2005;68:349-359. Review.
- Velagaleti GV, Bien-Willner GA, Northup JK, Lockhart LH, Hawkins JC, Jalal SM, Withers M, Lupski JR, Stankiewicz P. Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. *Am J Hum Genet*. 2005;76:652-662.
- Hill-Harfe KL, Kaplan L, Stalker HJ, Zori RT, Pop R, Scherer G, Wallace MR. Fine mapping of chromosome 17 translocation breakpoints > or = 900 Kb upstream of SOX9 in acampomelic campomelic dysplasia and a mild, familial skeletal dysplasia. *Am J Hum Genet*. 2005;76:663-671.
- Erdel M, Lane AH, Fresser F, Probst P, Utermann G, Scherer G. A new campomelic dysplasia translocation breakpoint maps 400 kb from SOX9. *Eur J Hum Genet* 2004;Suppl 12, 136. (Citado por Feuk et al. 2006-Referencia 6).
- Pop R, Conz C, Lindenberg KS, Blesson S, Schmalenberger B, Briault S, Pfeifer D, Scherer G. Screening of the 1 Mb SOX9 5' control re-

- gion by array CGH identifies a large deletion in a case of campomelic dysplasia with XY sex reversal. *J Med Genet.* 2004;41:e47.
24. David D, Cardoso J, Marques B, Marques R, Silva ED, Santos H, Boavida MG. Molecular characterization of a familial translocation implicates disruption of HDAC9 and possible position effect on TGFbeta2 in the pathogenesis of Peters' anomaly. *Genomics.* 2003;81:489-503.
 25. Marlin S, Blanchard S, Slim R, Lacombe D, Denoyelle F, Alessandri JL, Calzolari E, Drouin-Garraud V, Ferraz FG, Fourmaintraux A, Philip N, Toublanc JE, Petit C. Townes-Brocks syndrome: detection of a SALL1 mutation hot spot and evidence for a position effect in one patient. *Hum Mutat.* 1999;14:377-386.
 26. Rodriguez L, Zollino M, Mansilla E, Martinez-Fernandez ML, Perez P, Murdolo M, Martinez-Frias ML. The first 4p euchromatic variant in a healthy carrier having an unusual reproductive history. *Am J Med Genet A.* 2007;143:995-998.
 27. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) project. *Science* 2004;306:636-640.
 28. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007;447:799-816.
 29. Dermitzakis ET, Reymond A, Antonarakis SE. Conserved non-genic sequences - an unexpected feature of mammalian genomes. *Nat Rev Genet.* 2005;6:151-157. Review.
 30. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M. What is a gene, post-ENCODE? History and updated definition. *Genome Res.* 2007;17:669-681.
 31. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR. Genome-wide RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science.* 2007;316:1484-1488.
 32. Mattick JS. RNA regulation: a new genetics? *Nat Rev Genet.* 2004;5:316-23. Review.
 33. Mattick JS. The functional genomics of noncoding RNA. *Science.* 2005;309:1527-1528.
 34. Mattick JS, Makunin IV. Non-coding RNA. *Hum Mol Genet.* 2006;15 Spec No 1:R17-29. Review.
 35. Yang XJ. Multisite protein modification and intramolecular signaling. *Oncogene.* 2005;24:1653-1662. Review.
 36. Washietl S, Pedersen JS, Korbel JO, Stocsits C, Gruber AR, Hackermuller J, Hertel J, Lindemeyer M, Reiche K, Tanzer A, Ucla C, Wyss C, Antonarakis SE, Denoeud F, Lagarde J, Drenkow J, Kapranov P, Gingeras TR, Guigo R, Snyder M, Gerstein MB, Reymond A, Hofacker IL, Stadler PF. Structured RNAs in the ENCODE selected regions of the human genome. *Genome Res.* 2007;17:852-64.
 37. Tress ML, Martelli PL, Frankish A, Reeves GA, Wesselink JJ, Yeats C, Olason PL, Albrecht M, Hegyi H, Giorgetti A, Raimondo D, Lagarde J, Laskowski RA, Lopez G, Sadowski MI, Watson JD, Fariselli P, Rossi I, Nagy A, Kai W, Storling Z, Orsini M, Assenov Y, Blankenburg H, Huthmacher C, Ramirez F, Schlicker A, Denoeud F, Jones P, Kerrien S, Orchard S, Antonarakis SE, Reymond A, Birney E, Brunak S, Casadio R, Guigo R, Harrow J, Hermjakob H, Jones DT, Lengauer T, Orengo CA, Patthy L, Thornton JM, Tramontano A, Valencia A. The implications of alternative splicing in the ENCODE protein complement. *Proc Natl Acad Sci U S A.* 2007;104:5495-500.
 38. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet.* 2007;6:413-23.
 39. Gingeras TR. Origin of phenotypes: Genes and transcripts. *Genome Research* 2007;8:413-423.
 40. Prasanth KV, Spector DL. Eukariotic regulatory RNAs: An answer to the "genome complexity" conundrum. *Genes and Dev.* 2007;21:11-42.
 41. Li LC, Okino ST, Zhao H, Pookot D, Place RF, Urakami S, Enokida H, Dahiya R. Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci U S A.* 2006;103:17337-17342.
 42. Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE, Corey DR. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol.* 2007;3:166-173.
 43. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell.* 2007;130:89-100.
 44. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. *Nature.* 2007;448:83-86.